Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 36 (2004) 207–216
Protein Expression & Purification
www.elsevier.com/locate/yprep

Protein aggregation during overexpression limited by peptide extensions with large net negative charge

Yian-Biao Zhang, Jason Howitt,[1] Sean McCorkle, Paul Lawrence, Karen Springer, and Paul Freimuth*

Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA

Received 8 January 2004, and in revised form 20 April 2004
Available online 10 June 2004

Abstract

Folding of the human coxsackie and adenovirus receptor immunoglobulin (Ig) variable-type domain (CAR D1) during overexpression in the *Escherichia coli* cytoplasm was shown previously to be partially rescued by fusion to a 22-residue C-terminal peptide. Here, peptide sequence features required for solubilization and folding of CAR D1 and similar Ig variable-type domains from two other human membrane proteins were investigated. Peptide extensions with net negative charge > -6 fully solubilized CAR D1, and approximately half of the peptide-solubilized protein was correctly folded. The Ig variable-type domains from human A33 antigen and myelin P-zero proteins were only partially solubilized by peptide extensions with net charge of -12, however, and only the solubilized P-zero domain appeared to fold correctly whereas the A33 domain formed soluble microaggregates of misfolded protein. Our results suggest a model where the large net charge of peptide extensions increases electrostatic repulsion between nascent polypeptides. The resulting decrease in aggregation rate can enable some polypeptides to fold spontaneously into their native protein conformations. Analysis of the solubility and folding status of sets of structurally homologous proteins, such as the Ig variable-type domains described here, during overexpression could provide insights into how amino acid and gene sequences influence the efficiency of spontaneous protein folding.

Keywords: Fusion protein; Aggregation; Protein folding; Electrostatic repulsion

Protein production and characterization has been greatly facilitated by the development of systems for expression of proteins to high levels in homologous or heterologous cells [1,2], although many proteins are unable to fold correctly during overexpression and instead form insoluble aggregates. Under normal expression conditions, a subset of endogenous polypeptides can fold spontaneously [3], while folding of aggregation-prone polypeptides requires assistance from molecular chaperones, trans-acting factors that associate reversibly with nascent polypeptides to prevent aggregation during the protein folding process [4]. Polypeptide aggregation during overexpression therefore could result either from accumulation of high concentrations of folding intermediates in spontaneous folding pathways or from inefficient recognition or processing of polypeptide substrates by molecular chaperones. Culturing cells at reduced temperature or increasing cell chaperone capacity can in some cases increase the yields of correctly folded overexpressed proteins, but these approaches are not universally effective. Development of alternative strategies to minimize aggregation of nascent polypeptides therefore would enhance the utility of systems for protein overexpression.

Fusion of aggregation-prone polypeptides to carrier proteins frequently increases the solubility and in some cases promotes correct folding of polypeptides during overexpression in bacteria or other host cells. Soluble fusion partners commonly used for this purpose include the thioredoxin [5], NusA [6] and maltose-binding proteins [7] from *Escherichia coli*, and the glutathione

* Corresponding author. Fax: 1-631-344-3407.
E-mail address: freimuth@bnl.gov (P. Freimuth).
[1] Present address: Department of Biological Sciences, Biophysics Section, Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2BW, UK.

1046-5928/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2004.04.020

S-transferase (GST) protein from *Schistosoma japonicum* [8]. The mechanism by which these carrier proteins promote folding of target polypeptides is not fully understood and may not be universal. It was suggested, for example, that a hydrophobic cleft on the surface of maltose-binding protein might bind reversibly to exposed hydrophobic regions of nascent passenger polypeptides, steering the polypeptides toward their native conformation by a chaperone-like mechanism [9]. Another study suggested that decreased translation rates resulting from NusA-mediated transcriptional pausing might enable proper nucleation of passenger polypeptide folding during translation [6].

Here, we report that aggregation-prone Ig variable-type domains derived from three human membrane glycoproteins can be solubilized during overexpression in the *E. coli* cytoplasm by extension of the domain N- or C-terminus with highly acidic peptides. In two cases, substantial fractions of the peptide-solubilized domains were correctly folded whereas the third protein predominantly formed soluble microaggregates. Although much smaller than maltose-binding protein, GST, or the other fusion partners noted above, these peptides nonetheless appear to have similar enhancing effects on protein solubility and folding. Furthermore, potential folding mechanisms that would require activities associated with specific protein conformations, such as those proposed for maltose-binding protein and NusA, do not likely apply to these peptide extensions because peptides of this length (<40 residues) generally are unable to adopt stable tertiary conformations. Our results suggest that the peptide extensions described here may indirectly promote folding by increasing electrostatic repulsion between nascent polypeptides. The resulting delay in polypeptide aggregation would provide more time for proteins to fold spontaneously by chaperone-independent mechanisms. Amino acid and gene sequence characteristics of these three Ig-type domains that might influence their efficiencies of spontaneous, chaperone-independent folding are discussed.

Materials and methods

Protein expression

cDNA clones for human A33 antigen and human myelin P_0 protein were obtained from the Image Consortium [10] collection (Clone Nos. 2710753 and 3926008, respectively). Clone 2710753 is a partial A33 cDNA clone, in which correctly spliced exons 1–3 are fused to intron 3. This clone thus encodes the complete A33 D1 domain and N-terminal signal peptide, and a 10-residue C-terminal extension of the D1 domain specified by intron 3. Human CAR cDNA was obtained by reverse transcriptase-PCR of HeLa cell total RNA. DNA fragments encoding the D1 domains of CAR and A33, and the single extracellular domain of P_0 (P_0ex) were amplified from cDNA templates by PCR using *Pfu* Turbo polymerase (Stratagene, La Jolla, CA) and the primer sets shown in Table 1. Products were cloned into expression vector pET15b (Novagen, Madison, WI), or into derivatives of pET15b created to allow fusion of proteins to the solubility-enhancing tags shown in Table 2. Ligated DNAs were initially transformed into *E. coli* strain DH5α for characterization, and subsequently into strain BL21(DE3) for protein expression. The predicted nucleotide sequences of all expression constructs were confirmed by DNA sequence analysis. Bacterial strains were grown in LB medium supplemented with 150 mg/L penicillinG. Protein expression was induced in midlog phase cultures by addition of isopropyl-β-D-galactopyranoside (IPTG) to 1 mM.

DNA fragments encoding the peptide extensions listed in Table 2 were produced by PCR amplification of bacteriophage T7 genomic DNA or a subcloned region of the bacteriophage T3 DNA genome [11], using *Pfu* polymerase and primer sets (sequences not shown) to adapt fragment ends for ligation between the *Xho*I/*Blp*I or *Bam*HI/*Blp*I sites of pET15b. The amino acid sequence changes listed in Table 2 were introduced via mutagenic primers (sequences not shown), following established methods for PCR mutagenesis [12]. The predicted DNA sequences encoding all variant peptides were confirmed by DNA sequence analysis. DNA encoding the T7B9 peptide (Table 2) also was amplified with primers (not shown) which adapted it for cloning

Table 1
PCR primers for Ig domain protein expression

Protein[a]	Primers[b]
CAR D1	F-CTAGTGCATATGGGTATCACTACTCCT
CAA68868.1	R1-TCTGACTCGAGTTAACCTGAAGGCTTAACA
S21G[c]-A144	R2-TGACTTCTCGAGCGCACCT
A33 D1	F-CAGTCATATGATCTCTGTGGAAACTCC
Q99795	R1-CTAGCTCGAGTCATTTGGAGGGTGGCACGAGG ACCAACAGGCG
I22-K143	R2-CTAGCTCGAGTTTGGAGGGTGGCACGAGGACC AACAGGCG
P_0ex	F-GGAATTCCATATGATCGTGGTTTACACCG
AAH06491	R1-CAGACTCGAGTCACCTAGTTGGCACTTTTTC
I40-R163	R2-CAGACTCGAGCCTAGTTGGCACTTTTTC

[a] Protein database accession numbers are indicated under the protein name, and residue numbers that define the endpoints of the expressed protein fragments are indicated under the accession numbers.

[b] Forward primers are denoted F-. Primers denoted R1- are reverse primers that contain stop codons to terminate protein translation at the Ig domain C-terminal boundary. Stop codons were omitted from reverse sense primers denoted R2-, to allow fusion of Ig domain C-termini to T7-derived solubility tags.

[c] Amino acid substitution designed into PCR primer.

Table 2
Sequence and net charge of peptides fused to the C-terminus of CAR D1

Peptide	Sequence[a]	Net charge[b]
T7C	LEDPFQSGVMLGVASTVAASPEEASVTSTEETLTPAQEAARTRAANKARKEAELAAATAEQ	−6
T7B	LE[e]DP----------EEASVTSTEETLTPAQEAAR[d]TRAANKARKEAELAAATAEQ	−6
T7B1	LEDP----------EEASVTSTEETLTPAQEAARTR**PP**NKARKEAELAAATAEQ	−6
T7B2	LEDP----------EEASVTSTEETLTPAQEAARTR**GG**NKARKEAELAAATAEQ	−6
T7B3	LEDP----------EEASVTSTEETLTPAQEAARTRAANKARKEAEL---TAEQ	−6
T7B4	LEDP----------EEASVTSTEETLTPAQEAARTRAANKARKEAEL**E**A**E**TAEQ	−8
T7B5	LEDP----------EEASVTSTEETLTPAQEAARTRAA**A**KARKEAELAAATAEQ	−6
T7B6	LEDP----------EEASVTSTEETLTPAQEAARTR--KARKEAELAAATAEQ	−6
T7B7	LEDP----------EEASVTSTEETLTPAQEAARTRAANKARKEAELAA---	−5
T7B8	LEDP----------EEASVTSTEETLTPAQEAARTRAANKARKEAELAAA---	−5
T7B9	LEDP----------EEASVTSTEETLTPAQEAA**E**T**E**AANKARKEAEL**E**A**E**TAEQ	−12
T7B10	LEDP----------TPAQEAARTRAANKARKEAELAAATAEQ	−2
T7A	LEDP----------AANK[d]ARKEAELAAATAEQ	−3
T7A1	LEDP----------**ER**NK**E**RKEAELAAATAEQ	−4
T7A2	LEDP----------**ER**NK**E**RKEAEL**E**AATAEQ	−5
T7A3	LEDP----------**ER**NK**E**RKEAEL**E**A**E**TAEQ	−6
T3[c]	LEDPAVWEAGKVVAKGVGTADITATTSNGLIASCKVIVNAATS	−2
T3A	LEDPAVWEAGKVVAKGVGTADITATTSNGLIAS**SEEAD**NAATS	−6

[a] The first four residues (LEDP) of all peptides are encoded by a synthetic DNA linker that was introduced into the pET15b cloning vector (Novagen, Madison, WI) to introduce cleavage sites for the restriction endonucleases *Xho*I and *Bam*HI.

[b] The number of positive charges contributed by basic amino acids (lysine and arginine) and negative charges contributed by acidic amino acids (aspartate and glutamate) was summed. The COO^- group of the C-terminal residue was included in the calculation.

[c] Only the C-terminal 39 amino acids of the 89-residue frameshifted T3 sequence were tested, to approximate the length of the T7B peptide.

[d] Site of cleavage by trypsin.

[e] Site of cleavage by V8 protease.

between the *Nco*I/*Nde*I sites of pET15b, to allow extension of protein N-termini with the T7B9 peptide.

Protein solubility test

Cells in 1.5 ml of induced culture were collected by centrifugation, resuspended in 0.3 ml STE buffer (10 mM Tris, 1 mM EDTA, and 100 mM NaCl, pH 8), and sonicated for three bursts of ~1-s duration at maximum output using a Misonix model 550 Sonic Dismembrator (Fisher Scientific, Pittsburg, PA) equipped with a microprobe tip. After centrifugation for 5 min at 16,000*g*, supernatants (soluble fraction) were transferred to fresh tubes, and pellets (insoluble fraction) were dispersed in 0.3 ml STE buffer. Equal volumes of each fraction were electrophoresed in sodium dodecyl sulfate–polyacrylamide gels to assess protein content.

Protein purification

CAR D1-T7B fusion protein was purified from cell lysates by anion exchange chromatography on Whatman DE52 cellulose (Fisher Scientific, Pittsburg, PA) followed by affinity chromatography on immobilized recombinant knob domain (knob) of the adenovirus-12 fiber protein [13]. Knob protein was coupled to cyanogen bromide-activated Sepharose beads (Sigma, St. Louis, MO) according to instructions provided by the supplier. CAR D1 was eluted from the knob affinity column in 100 mM glycine buffer, pH 2.5, and immediately adjusted to neutral pH by addition of 0.1 volume of 1 M Tris buffer, pH 8.5. Purified CAR D1 was dialyzed against phosphate-buffered saline (PBS) before further analysis by protease digestion. Conditions for electrophoresis of CAR D1 and knob–CAR D1 complexes in non-denaturing polyacrylamide gels were previously described [14].

Soluble 6-His tagged A33 D1 and P_0ex fusion proteins were purified by metal-affinity chromatography on nickel–NTA agarose resin (Qiagen, Valencia, CA) according to methods recommended by the supplier. Proteins eluting from the column in 250 mM imidazole were dialyzed against size exclusion chromatography buffer (10 mM Tris–HCl, 100 mM NaCl, pH 8) before storage at −80 °C or further purification by gel filtration. Gel filtration analysis was performed on a Sephacryl S-200 column using an AKTA Prime low pressure liquid chromatograph (Amersham–Pharmacia Biotech, Piscataway, NJ) and a flow rate of 0.5 ml/min. The column was calibrated using a 12–200 kDa molecular weight range marker kit (Sigma, St. Louis, MO).

Insoluble A33-T7B9 fusion protein and unmodified A33 protein in the pellet fraction of bacterial cell lysates were dissolved in 6 M guanidine–HCl and purified by metal-affinity chromatography on Ni–NTA–agarose beads (Qiagen, Valencia, CA). The proteins were eluted in 8 M urea and dialyzed against 10 mM sodium phosphate buffer, pH 7.2. After centrifugation for 10 min at 16,000*g*, proteins in the supernatant were applied to a carbon grid for examination in the scanning transmission electron microscope at Brookhaven National Laboratory. Masses of unstained proteins were determined

using the program PC mass (J. Wall, personal communication).

Limited proteolysis

Trypsin or V8 protease was added (10 μg/ml) to protein solutions, and the mixtures were incubated at 37 °C. At the specified time points, portions of the reactions were withdrawn, mixed with electrophoresis sample buffer, and boiled for 5 min to terminate digestion. Samples were then analyzed by electrophoresis in 12% polyacrylamide gels containing sodium dodecyl sulfate [15] or in Tris–Tricine-buffered 10% polyacrylamide gels [16]. Masses of stable trypsin digestion products of P_0ex were determined in a Voyager DE mass spectrometer (PerSeptive Biosystems, Foster City, CA) using sinapinic acid matrix.

Calculation of protein integrated net charge

A PERL program was written to calculate integrated charge for input amino acid sequences. It accumulates a charge of +1 for the first methionine and for each lysine and arginine in the sequence, a charge of −1 for each glutamate and aspartate, and 0 (neutral) for all other residues. The running total is output at each amino acid position. For ease of use, a PHP web interface was then added, which accepts text sequences (i.e., cut-and-paste input) and provides graphical display (via the open source program GNUplot) as well as tabular output of the integrated charge function.

Results

Rescue of CAR D1 folding

Consistent with our earlier report [13], the coxsackie and adenovirus receptor [17,18] amino-terminal domain (CAR D1) aggregated completely during overexpression in the *E. coli* cytoplasm, but was partially solubilized when the domain C-terminus was extended by a 22-residue C-terminal peptide encoded by plasmid expression vector sequences (Fig. 1A, peptide T7A). A substantial amount of the peptide-solubilized CAR D1 previously was shown to be correctly folded and biologically active [13,19]. CAR D1 is 120 amino acids in length and folds into a disulfide-bonded β-sandwich conformation characteristic of immunoglobulin variable-type (IgV) domains [19,20]. The 22-residue peptide fused to CAR D1 corresponds to the last 18 residues of a 57-residue C-terminal extension present on the bacteriophage T7 minor capsid protein 10B [21], plus four additional residues (LEDP) encoded by an *Xho*I–*Bam*HI DNA linker in the pET15b (Novagen, Madison, WI) expression vector (Table 2, peptide T7A). CAR D1 was completely solubilized when fused to either the entire 57-residue T7-derived peptide or to the last 40 residues of this peptide (Fig. 1A and Table 2, peptides T7C and T7B, respectively). By contrast, CAR D1 was not solubilized when fused to the last 40 residues of the C-terminal extension present on the analogous minor capsid protein of bacteriophage T3 (Fig. 1A and Table 2, peptide T3). The T3-derived peptide has no significant sequence homology to the corresponding T7 peptide [11].

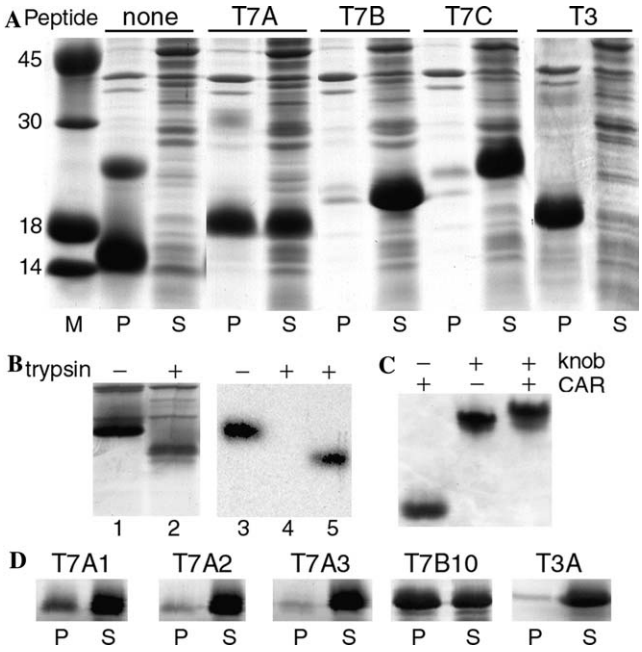

Fig. 1. Effect of C-terminal extensions on CAR D1 solubility and folding. (A) SDS–PAGE analysis of the insoluble and soluble fractions of cell lysates (lanes P and S, respectively) of *E. coli* strain BL21-DE3 expressing CAR D1 either without a C-terminal extension (none) or fused to C-terminal peptides T7A, T7B, T7C, or T3. Sequences of C-terminal peptides are shown in Table 2. Cell suspensions in STE buffer (10 mM Tris–HCl, 100 mM NaCl, and 1 mM ethylenediaminetetraacetic acid, pH 8.0) were sonicated and then centrifuged for 5 min at 15,000*g* to separate insoluble and soluble fractions. Molecular weights in kiloDalton of protein standards loaded in lane M are indicated. (B) SDS–PAGE and Western immunoblot analysis of the soluble fraction of an *E. coli* lysate containing CAR D1-T7B before and after digestion for 20 min with 10 μg/ml trypsin. Lanes 1 and 2 are a Coomassie blue-stained gel. Lanes 3–5 are an immunoblot probed with rabbit antisera raised against either synthetic peptide T7A (lanes 3 and 4) or recombinant CAR D1 (lane 5). (C) Non-denaturing polyacrylamide gel electrophoresis of trypsin-treated CAR D1-T7B (1 μg) alone or after mixing with a molar excess of knob domain from the adenovirus-2 fiber protein (2 μg). Note that CAR D1 quantitatively assembled into complexes with Ad2 knob. (D) SDS–PAGE analysis of the insoluble and soluble fractions (lanes P and S, respectively) of lysates of *E. coli* strains expressing CAR D1 fused to C-terminal peptides T7A1, T7A2, T7A3, T7B10, or T3A. Sequences of C-terminal peptides are shown in Table 2.

As observed with the CAR D1-T7A fusion protein [13], the T7B C-terminal peptide extension was more sensitive than the CAR D1 domain to protease digestion (Fig. 1B; trypsin and V8 cleavage sites in T7A and T7B peptides are indicated in Table 2) and the resulting

protease-stable CAR D1 fragment remained in solution and could form complexes with the recombinant knob domain of the adenovirus fiber protein (Fig. 1C), indicating that the T7B extension also promotes correct folding of CAR D1. Size exclusion chromatography showed that CAR D1-T7B protein in crude cell sonicates had a heterogeneous aggregation state (Figs. 2A and B). CAR D1-T7B protein eluting in the column excluded volume was trypsin-sensitive and therefore misfolded and microaggregated, whereas the fraction of CAR D1-T7B eluting in the included volume was trypsin-resistant (Fig. 2C). Approximately half of the total soluble CAR D1-T7B protein was in the trypsin-resistant conformation.

Mechanism of peptide-mediated solubilization and folding of CAR D1

Sequence characteristics of the T7B peptide necessary for enhanced solubility and folding of CAR D1 were investigated by mutagenesis. The T7B peptide has a predicted [22] amphipathic α-helix extending from Leu12 to Ala36 (see Table 2 for numbering convention) that was either shortened or reduced in amphipathic character in peptide variants T7B1–T7B4 (Table 2). The T7B peptide also contains sequence motifs similar to chaperone recognition motifs in the *E. coli* ssrA peptide [23,24] that were disrupted in peptide variants T7B5–T7B8 (Table 2). However, peptide variants T7B1–T7B8 all fully solubilized CAR D1 (data not shown).

C-terminal peptides that fully solubilized CAR D1 had net charges in the range of −5 to −8 (peptides T7C, T7B, and T7B1–T7B8), whereas the partially active T7A peptide and the inactive T3-derived peptide had net charges of −3 and −2, respectively (Table 2). Additional peptide variants were constructed to further examine the relationship between peptide net negative charge and ability to solubilize CAR D1. CAR D1 solubility progressively increased when fused to T7A variants with increasing net charges of −4, −5, or −6 (Fig. 1D and

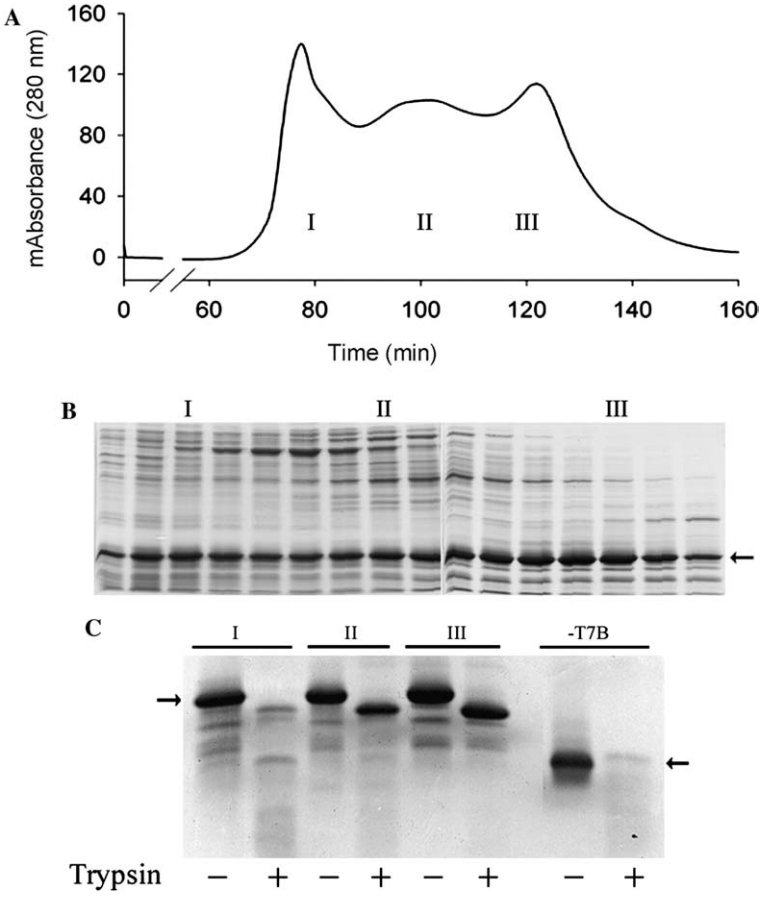

Fig. 2. Analysis of CAR-T7B folding and aggregation state. (A) Size exclusion chromatography of cleared lysate of *E. coli* cells expressing CAR D1-T7B. Cells were disrupted with BugBuster reagent (Novagen), lysate was clarified by centrifugation for 1 h at 100,000*g* and then chromatographed on a Sephacryl S-200 column. The column was calibrated as indicated in Materials and methods. (B) SDS–PAGE of protein content in eluant fractions from chromatography in (A). Representative fractions from the void volume peak and from two regions of the included volume are designated I, II, and III, respectively. (C) SDS–PAGE of protein in fractions I, II, and III after digestion for 30 min at 37 °C with 10 μg/ml trypsin. Insoluble aggregates of unmodified CAR D1 were dispersed in buffer and digested in parallel with trypsin for comparison (loaded in lanes marked −T7B). Trypsin only partially removes the T7B extension (see Table 2, Footnote d), accounting for the slower mobility of trypsin-treated species II and III compared to unmodified CAR D1. Arrows in (B and C) indicate position of CAR D1-T7B and unmodified CAR D1 before digestion with trypsin.

Table 2, peptides T7A1–T7A3). Consistent with this trend, CAR D1 solubility decreased when fused to a T7B variant with net charge of −2 (Fig. 1D and Table 2, peptide T7B10), and its solubility increased when fused to a T3 variant with net charge of −6 (Fig. 1D and Table 2, peptide T3A). Together these results indicate that excess negative charge is the key feature required for peptide extensions to increase the solubility of CAR D1. Peptide length and amino acid sequence context appear to be of minor importance for this activity.

Effect of peptides on solubility and folding of other Ig variable-type domains

We next examined whether charged peptide extensions can also enhance the solubility and folding of other IgV-type domains that are similar in amino acid sequence to CAR D1 during overexpression in the *E. coli* cytoplasm. The amino-terminal domain of human A33 antigen [25] was identified by BLAST homology search [26] as the protein domain most closely related in amino acid sequence to CAR D1 (30% identical and 53% similar). Amino acid sequence alignment predicted that A33 D1 also has an IgV-type fold [27–29]. The extracellular domain of myelin P_0 protein (P_0ex) from shark is the next protein domain most similar to CAR D1 (29% identical and 45% similar). The X-ray structure of rat P_0ex [30] indicated that the P_0ex domain indeed folds into an IgV-type conformation. cDNA fragments encoding the IgV-type domains of human A33 antigen and human P_0ex (human P_0ex is 27% identical and 40% similar to CAR D1) were cloned into the expression vector pET15b.

A33 D1 was completely insoluble when expressed in *E. coli* without a C-terminal peptide extension (Fig. 3A, inset), but unlike CAR D1 it remained completely insoluble at 37 °C and was only partially soluble at 25 °C when fused to the C-terminal T7B peptide (not shown). When A33 D1 was fused to the T7B9 variant peptide with net charge of −12 (Table 2), the protein solubility increased substantially (Fig. 3A, inset), further supporting the conclusion that the ability of peptide extensions to solubilize proteins increases with the magnitude of the peptide net negative charge. After purification, soluble A33 D1-T7B9 fusion protein migrated as several distinct species during electrophoresis in polyacrylamide gels under non-denaturing conditions (Fig. 3A, inset lane ND) and had a heterogeneous aggregation state as revealed by size exclusion chromatography (Figs. 3A and B). In contrast to CAR D1-T7B, the IgV-type domain and the peptide extension of the A33 D1-T7B9 fusion protein were equally sensitive to digestion by trypsin (Fig. 3A, inset), suggesting that the A33 IgV-type domain was not correctly folded.

Refolding of A33 D1 in vitro was examined to further assess the ability of this domain to fold spontaneously. Aggregated A33 D1 and A33 D1-T7B9 polypeptides

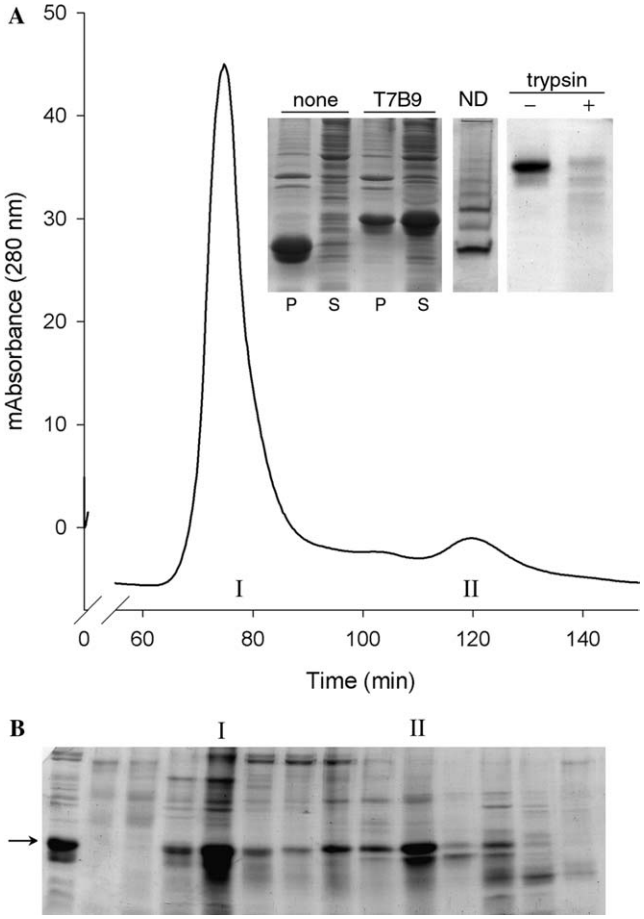

Fig. 3. Aggregation state of A33 D1-T7B9 fusion protein. (A) Size exclusion chromatography of purified 6-His-A33 D1-T7B9 protein on a Sephacryl S-200. Inset: SDS–PAGE of the insoluble and soluble fractions (lanes P and S) of lysates of *E. coli* expressing A33 D1 without a peptide extension (none) or with the C-terminal T7B9 extension (T7B9); non-denaturing polyacrylamide gel electrophoresis of A33 D1-T7B9 protein (ND); SDS–PAGE of A33 D1-T7B9 before and after digestion with trypsin (−, +). (B) SDS–PAGE analysis of protein content of column fractions in (A). Peak fractions of A33 D1-T7B9 in the excluded and included volumes are indicated by I and II, respectively. A sample of purified 6-His-A33 D1-T7B9 before loading onto the Sephacryl S-200 column was loaded in the first lane on the left. Arrow indicates the position of A33-D1-T7B9 in the gel.

were isolated from *E. coli* cells, dissolved in 6 M guanidine–hydrochloride, and purified by metal-affinity chromatography under denaturing conditions. The bulk of unmodified A33 D1 precipitated during removal of denaturant by dialysis, whereas the majority of A33 D1-T7B9 fusion protein remained in solution, even at relatively high protein concentrations (5 mg/ml). Most of the soluble, renatured A33 D1-T7B9 was microaggregated, however, with masses ranging up to 5 MDa as measured by scanning transmission electron microscopy (Fig. 4A). These soluble microaggregates failed to migrate in polyacrylamide gels during electrophoresis under non-denaturing conditions, but did migrate as a single band to the expected position for monomeric A33 D1 after brief heat

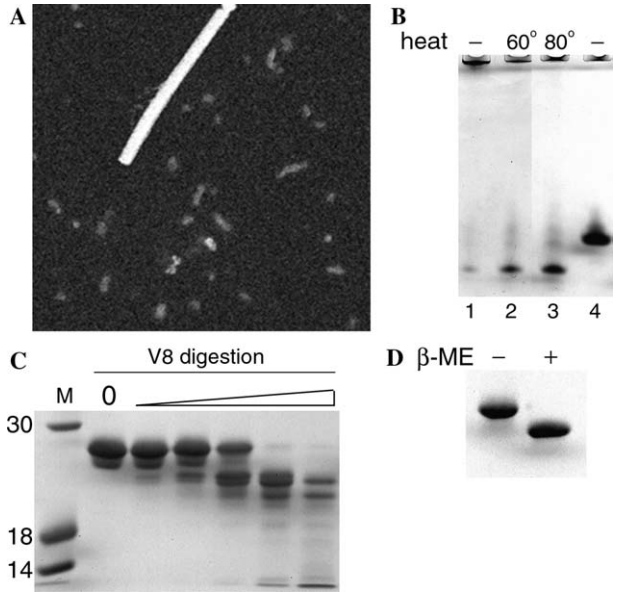

Fig. 4. Effect of T7B9 C-terminal peptide on folding of A33 D1 in vitro. (A) Scanning transmission electron microscopy (STEM) of renatured A33 D1-T7B9 fusion protein before heating. Masses of protein aggregates were measured relative to tobacco mosaic virus (rod-shaped object) which was mixed with the protein sample to provide an internal mass standard. (B) Non-denaturing PAGE of renatured A33 D1-T7B9 fusion protein before (lane 1) or after heating for 5 min at 60 or 80 °C (lanes 2–3, respectively). A sample of CAR D1 was loaded in lane 4 for comparison. (C) SDS–PAGE of time course of V8 protease digestion of renatured, heated A33 D1-T7B9 fusion. Triangle denotes increasing times (from 5 to 30 min) of digestion of A33 D1-T7B9 protein with 10 μg/ml V8 protease at 37 °C. (D) SDS–PAGE of renatured, heated A33 D1-T7B9 fusion protein under reducing and non-reducing conditions (±5% β-mercaptoethanol in sample buffer).

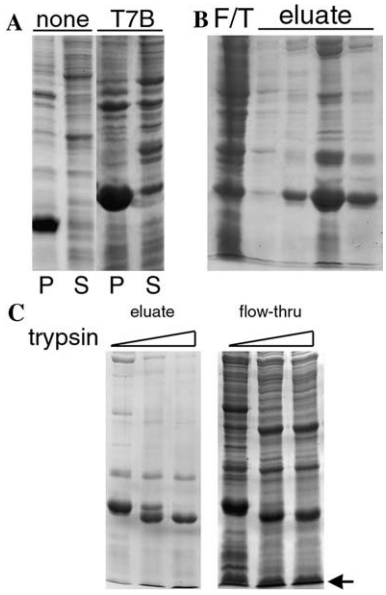

Fig. 5. Effect of T7B C-terminal extension on human P_0ex solubility and folding. (A) SDS–PAGE of the insoluble and soluble fractions (lanes marked P and S, respectively) of sonicates of *E. coli* cells expressing 6-His-P_0ex without a C-terminal extension (none) or fused to the T7B C-terminal extension (T7B). (B) SDS–PAGE analysis of the flow-through (F/T) and eluted fractions (eluate) of metal-affinity chromatography of 6-His-P_0ex-T7B protein. (C) SDS–PAGE analysis of time course of trypsin digestion of proteins in metal-affinity column eluate or flow-through fractions. Triangle denotes increasing time of digestion. Arrow indicates position of 7.8 kDa C-terminal fragment of P_0ex generated by trypsin digestion after lysine 55.

treatment at >60 °C (Fig. 4B). The C-terminal T7B9 extension of the heated protein was more sensitive than the A33 D1 domain to protease digestion (Fig. 4C). Finally, the electrophoretic mobility of the heated, renatured A33 D1-T7B9 protein in SDS–polyacrylamide gels increased when reducing agent was omitted from the sample buffer (Fig. 4D), indicating that the intra-domain disulfide bond had formed and consequently that A33 D1 was correctly folded.

The human P_0ex domain also was completely insoluble in the *E. coli* cytoplasm when expressed without a C-terminal extension, and like A33 D1, it was only partially solubilized by fusion to the C-terminal T7B extension (Fig. 5A). Soluble 6-His-P_0ex-T7B protein did not bind quantitatively to nickel affinity resin and the bound fraction eluted with several contaminating host cell proteins (Fig. 5B) that could not be removed by a second passage over the metal-affinity column. However, P_0ex protein in both the bound and flow-through fractions was equally resistant to protease digestion compared to the T7B extension (Fig. 5C). Furthermore, P_0ex protein in both the bound and flow-through fractions was cleaved by trypsin preferentially at lysine 55 (Fig. 5C), which is conserved in rat P_0ex and highly solvent-exposed in the rat P_0ex crystal structure [30]. The hypersensitivity of lysine 55 of human P_0ex to digestion by trypsin therefore is evidence that this domain may be correctly folded.

The effect of an N-terminal peptide extension on human P_0ex solubility and folding also was examined. When the T7B9 peptide (-12 charge) was fused to the P_0ex N-terminus, about half of the fusion protein was in the soluble fraction of cell sonicates (Fig. 6A). Digestion of soluble T7B9-P_0ex with trypsin generated a stable peptide of ~8 kDa (Fig. 6B), that was determined by mass spectrometry to correspond to a 7.8 kDa C-terminal fragment resulting from cleavage at lysine residue 55 (see above). Taken together, our results suggest that the soluble P_0ex-T7B and T7B9-P_0ex proteins both are correctly folded. The atypical behavior of P_0ex during chromatography may reflect the propensity of P_0ex proteins to interact with membrane phospholipids through the solvent-exposed hydrophobic side chains of two conserved tryptophan residues [30,31]. Adhesive interactions between these solvent-exposed hydrophobic side chains also might contribute to the mechanism of reversible gelation of recombinant rat and shark P_0ex proteins that was previously reported [30].

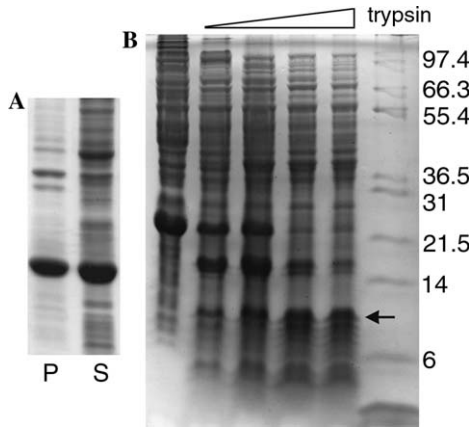

Fig. 6. Effect of T7B9 N-terminal extension on solubility and folding of human P_0ex. (A) SDS–PAGE of the insoluble and soluble fractions (lanes P and S, respectively) of a sonicate of *E. coli* cells expressing the T7B9-P_0ex fusion protein. (B) High resolution SDS–PAGE (Tris–Tricine gel) analysis of the time course of digestion of T7B9-P_0ex with trypsin. Triangle denotes increasing time of digestion with 10 μg/ml trypsin at room temperature. Arrow indicates position of 7.8 kDa C-terminal fragment of P_0ex generated by trypsin digestion at lysine 55.

Discussion

Short peptides such as those used as fusion partners in this study generally are unable to adopt stable tertiary structures, therefore the enhanced folding of CAR D1 and P_0ex conferred by these peptides most likely results from a simple mechanism that is independent of peptide conformation. The highly charged peptide extensions could increase electrostatic repulsion between nascent polypeptides, limiting polypeptide aggregation and prolonging the time that nascent polypeptides remain in productive folding pathways. This mechanism could account for both the correct folding of substantial fractions of the CAR D1 and P_0ex polypeptides and the formation of soluble microaggregates of the A33 D1 polypeptide. According to this model, the peptide extensions do not directly promote folding of associated polypeptides but rather delay polypeptide aggregation long enough to allow polypeptides to proceed further along spontaneous folding pathways [3] or to be recognized as substrates for assisted folding by host cell molecular chaperones.

NusA, maltose-binding protein, and some other intact proteins that often are used as fusion partners have large net negative charges and therefore also may indirectly promote folding of passenger proteins by limiting polypeptide aggregation, rather than actively promoting the folding of passenger polypeptides by a chaperone-like mechanism as previously suggested [9]. An indirect mechanism is consistent with the observation that maltose-binding protein mutants with slow folding kinetics lose the ability to promote folding of aggregation-prone C-terminal passenger proteins [32]. Long-lived folding intermediates of maltose-binding protein mutants likely would have exposed hydrophobic surfaces that could interact directly with nascent passenger polypeptides and inhibit their spontaneous folding. Rapid folding of wild-type maltose-binding protein into its native conformation would minimize hydrophobic interactions with nascent passenger polypeptides and thus minimize interference with passenger polypeptide folding.

Peptide-solubilized A33 D1 did not fold correctly in vivo but instead formed soluble microaggregates. Similar microaggregates or "soluble inclusion bodies" have been observed when other aggregation-prone proteins were fused to large carrier proteins such as the maltose-binding protein [33–37]. A33 D1 also was difficult to refold in vitro, requiring sequential chemical and heat denaturation to attain a native-like conformation. Together these results suggest there is a large free energy barrier to correct folding of the A33 D1 polypeptide by a spontaneous, chaperone-independent mechanism. In native, full-length A33 protein, the stability of the D1 domain may be influenced by the adjacent extracellular D2 domain, which is not present in our expression construct, or from posttranslational modifications that do not occur in the bacterial expression system. Full-length CAR also has two extracellular domains, however the occurrence of mRNA splice variants that encode the CAR D1 domain alone [18,38,39] supports the conclusion that the D1 domain has the capacity to fold into its native conformation in the absence of the adjacent D2 domain. By contrast, A33 mRNA splice variants that might encode the A33 D1 domain alone have not yet been reported (however see the description of the A33 cDNA clone in the Materials and methods section).

An earlier study found that the ability of overexpressed polypeptides to fold correctly was correlated with polypeptide net negative charge and content of amino acids that are commonly found in β-turns [40]. Consistent with this prediction, the ability of CAR D1, P_0ex, and A33 D1 to fold correctly also varied according to polypeptide net charge (−3, −4, and −1, respectively), although the differences in net charge between these polypeptides are relatively small. In contrast to the small differences in overall net charge, there is a large difference between these three IgV-type domains in the distribution of charged residues across the polypeptide chain. The CAR D1 polypeptide gains a large net negative charge early during its synthesis, whereas the A33 D1 polypeptide acquires a substantial net charge only after synthesis of the first ~100 residues. The large net negative charge of the CAR D1 polypeptide could produce an intramolecular repulsive force that destabilizes early folding intermediates, effectively lowering the free energy barrier to the final native conformation. Absence of strong intramolecular repulsion in the A33 D1 polypeptide might stabilize and kinetically trap early folding

intermediates that ultimately aggregate after release from the ribosome. Intramolecular electrostatic repulsion on protein surfaces resulting from clustered positively or negatively charged residues has been shown to destabilize native protein conformations [41,42].

Codon usage is another factor that potentially could influence the efficiency of spontaneous folding of these three IgV-type domains. The CAR D1 and P_0ex genes have a large number (40 and 28, respectively) of codons with strong or weak codon:anticodon interaction energies [43], whereas the A33 D1 gene has only 19 such codons. Codons with strong or weak codon:anticodon interaction energies decrease the rate of translation [43], possibly providing more time for the nascent polypeptide to nucleate folding cotranslationally. Further investigation of the solubility and folding of sets of structurally homologous proteins during overexpression would likely provide insights into how amino acid and gene sequences influence spontaneous protein folding.

Acknowledgments

We thank I. Molineux for providing plasmid DNA encoding the T3 10B capsid protein, J. Wall and members of the Brookhaven National Lab STEM group for mass analysis of protein aggregates, and C.W. Anderson, F.W. Studier, and C. Brakel for many helpful discussions throughout this project. This research was supported by US D.O.E. Laboratory-Directed Research and Development Grant 00-45 at Brookhaven National Laboratory and by NIH Grant AI36251 (both to P.F.).

References

[1] L.K. Miller, Baculoviruses for foreign gene expression in insect cells, Biotechnology 10 (1988) 457–465.
[2] F.W. Studier, A.H. Rosenberg, J.J. Dunn, J.W. Dubendorff, Use of T7 RNA polymerase to direct expression of cloned genes, Methods Enzymol. 185 (1990) 60–89.
[3] C.B. Anfinsen, Principles that govern the folding of protein chains, Science 181 (1973) 223–230.
[4] F.U. Hartl, M. Hayer-Hartl, Molecular chaperones in the cytosol: from nascent chain to folded protein, Science 295 (2002) 1852–1858.
[5] E.R. LaVallie, E.A. DiBlasio, S. Kovacic, K.L. Grant, P.F. Schendel, J.M. McCoy, A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm, Biotechnology (NY) 11 (1993) 187–193.
[6] G.D. Davis, C. Elisee, D.M. Newham, R.G. Harrison, New fusion protein systems designed to give soluble expression in *Escherichia coli*, Biotechnol. Bioeng. 65 (1999) 382–388.
[7] C. di Guan, P. Li, P.D. Riggs, H. Inouye, Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein, Gene 67 (1988) 21–30.
[8] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase, Gene 67 (1988) 31–40.
[9] R.B. Kapust, D.S. Waugh, *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, Protein Sci. 8 (1999) 1668–1674.
[10] G. Lennon, C. Auffray, M. Polymeropoulos, M.B. Soares, The I.M.A.G.E. consortium: an integrated molecular analysis of genomes and their expression, Genomics 33 (1996) 151–152.
[11] J.P. Condreay, S.E. Wright, I.J. Molineux, Nucleotide sequence and complementation studies of the gene 10 region of bacteriophage T3, J. Mol. Biol. 207 (1989) 555–561.
[12] G. Sarkar, S.S. Sommer, The megaprimer method of site-directed mutagenesis, Biotechniques 8 (1990) 404–407.
[13] P. Freimuth, K. Springer, C. Berard, J. Hainfeld, M. Bewley, J. Flanagan, Coxsackievirus and adenovirus receptor amino-terminal immunoglobulin V-related domain binds adenovirus type 2 and fiber knob from adenovirus type 12, J. Virol. 73 (1999) 1392–1398.
[14] M.J. van Raaij, N. Louis, J. Chroboczek, S. Cusack, Structure of the human adenovirus serotype 2 fiber head domain at 1.5 Å resolution, Virology 262 (1999) 333–343.
[15] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[16] H. Schagger, G. von Jagow, Tricine–sodium dodecyl sulfate–polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, Anal. Biochem. 166 (1987) 368–379.
[17] J.M. Bergelson, J.A. Cunningham, G. Droguett, E.A. Kurt-Jones, A. Krithivas, J.S. Hong, M.S. Horwitz, R.L. Crowell, R.W. Finberg, Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5, Science 275 (1997) 1320–1323.
[18] R.P. Tomko, R. Xu, L. Philipson, HCAR and MCAR: the human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses, Proc. Natl. Acad. Sci. USA 94 (1997) 3352–3356.
[19] M.C. Bewley, K. Springer, Y.B. Zhang, P. Freimuth, J.M. Flanagan, Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR, Science 286 (1999) 1579–1583.
[20] M.J. van Raaij, E. Chouin, H. van der Zandt, J.M. Bergelson, S. Cusack, Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 Å resolution, Struct. Fold. Des. 8 (2000) 1147–1155.
[21] B.G. Condron, J.F. Atkins, R.F. Gesteland, Frameshifting in gene 10 of bacteriophage T7, J. Bacteriol. 173 (1991) 6998–7003.
[22] J. Garnier, J.F. Gibrat, B. Robson, GOR method for predicting protein secondary structure from amino acid sequence, Methods Enzymol. 266 (1996) 540–553.
[23] J.M. Flynn, I. Levchenko, M. Seidel, S.H. Wickner, R.T. Sauer, T.A. Baker, Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis, Proc. Natl. Acad. Sci. USA 98 (2001) 10584–10589.
[24] I. Levchenko, M. Seidel, R.T. Sauer, T.A. Baker, A specificity-enhancing factor for the ClpXP degradation machine, Science 289 (2000) 2354–2356.
[25] J.K. Heath, S.J. White, C.N. Johnstone, B. Catimel, R.J. Simpson, R.L. Moritz, G.F. Tu, H. Ji, R.H. Whitehead, L.C. Groenen, A.M. Scott, G. Ritter, L. Cohen, S. Welt, L.J. Old, E.C. Nice, A.W. Burgess, The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily, Proc. Natl. Acad. Sci. USA 94 (1997) 469–474.
[26] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[27] I. Chretien, A. Marcuz, M. Courtet, K. Katevuo, O. Vainio, J.K. Heath, S.J. White, L. Du Pasquier, CTX, a *Xenopus* thymocyte receptor, defines a molecular family conserved throughout vertebrates, Eur. J. Immunol. 28 (1998) 4094–4104.
[28] Y. Harpaz, C. Chothia, Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains, J. Mol. Biol. 238 (1994) 528–539.

[29] A.F. Williams, A.N. Barclay, The immunoglobulin superfamily—domains for cell surface recognition, Annu. Rev. Immunol. 6 (1988) 381–405.
[30] L. Shapiro, J.P. Doyle, P. Hensley, D.R. Colman, W.A. Hendrickson, Crystal structure of the extracellular domain from P0, the major structural protein of peripheral nerve myelin, Neuron 17 (1996) 435–449.
[31] C.A. Wells, R.A. Saavedra, H. Inouye, D.A. Kirschner, Myelin P0-glycoprotein: predicted structure and interactions of extracellular domain, J. Neurochem. 61 (1993) 1987–1995.
[32] J.D. Fox, R.B. Kapust, D.S. Waugh, Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins, Protein Sci. 10 (2001) 622–630.
[33] Y. Nomine, T. Ristriani, C. Laurent, J.F. Lefevre, E. Weiss, G. Trave, Formation of soluble inclusion bodies by hpv e6 oncoprotein fused to maltose-binding protein, Protein Expr. Purif. 23 (2001) 22–32.
[34] J.M. Louis, R.A. McDonald, N.T. Nashed, E.M. Wondrak, D.M. Jerina, S. Oroszlan, P.T. Mora, Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in *Escherichia coli*, Eur. J. Biochem. 199 (1991) 361–369.
[35] H.K. Lorenzo, D. Farber, V. Germain, O. Acuto, P.M. Alzari, The MBP fusion protein restores the activity of the first phosphatase domain of CD45, FEBS Lett. 411 (1997) 231–235.
[36] V.M. Saavedra-Alanis, P. Rysavy, L.E. Rosenberg, F. Kalousek, Rat liver mitochondrial processing peptidase. Both alpha- and beta-subunits are required for activity, J. Biol. Chem. 269 (1994) 9284–9288.
[37] D. Sachdev, J.M. Chirgwin, Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein, J. Protein Chem. 18 (1999) 127–136.
[38] H. Fechner, A. Haack, H. Wang, X. Wang, K. Eizema, M. Pauschinger, R. Schoemaker, R. Veghel, A. Houtsmuller, H.P. Schultheiss, J. Lamers, W. Poller, Expression of coxsackie adenovirus receptor and alphav-integrin does not correlate with adenovector targeting in vivo indicating anatomical vector barriers, Gene Ther. 6 (1999) 1520–1535.
[39] I. Thoelen, C. Magnusson, S. Tagerud, C. Polacek, M. Lindberg, M. Van Ranst, Identification of alternative splice products encoded by the human coxsackie-adenovirus receptor gene, Biochem. Biophys. Res. Commun. 287 (2001) 216–222.
[40] D.L. Wilkinson, R.G. Harrison, Predicting the solubility of recombinant proteins in *Escherichia coli*, Biotechnology (NY) 9 (1991) 443–448.
[41] C. Nishimura, V.N. Uversky, A.L. Fink, Effect of salts on the stability and folding of staphylococcal nuclease, Biochemistry 40 (2001) 2113–2128.
[42] D.F. Wyss, J.S. Choi, J. Li, M.H. Knoppers, K.J. Willis, A.R. Arulanandam, A. Smolyar, E.L. Reinherz, G. Wagner, Conformation and function of the N-linked glycan in the adhesion domain of human CD2, Science 269 (1995) 1273–1278.
[43] H. Grosjean, W. Fiers, Preferential codon usage in prokaryotic genes: the optimal codon–anticodon interaction energy and the selective codon usage in efficiently expressed genes, Gene 18 (1982) 199–209.